VOL. 15, NO. 2, 2022
ISSN 1936-878X/$36.00

iMAIL

Deep Learning for Automatic Calcium Scoring in Population-Based Cardiovascular Screening

The detection and quantification of coronary artery calcifications (CAC) with noncontrast electrocardiography-triggered computed tomography (CT) can predict future cardiovascular disease in asymptomatic individuals (1,2). In a CAC screening setting, high volumes of CT scans are generated and need to be scored. Scoring of CAC is a highly repetitive and labor-intensive task for radiologists, but multiple artificial intelligence applications have been shown to be able to automate this task. Nevertheless, studies that evaluated the performance of artificial intelligence applications for automatic CAC scoring (mainly) included clinically indicated CT scans varying from dedicated CAC CT and chest CT to positron emission tomography CT as well as poorly defined retrospective study populations with varying CAC distributions (3,4).

The objective of the current study was to evaluate the performance of deep learning-based software for automatic coronary calcium scoring in a screening setting.

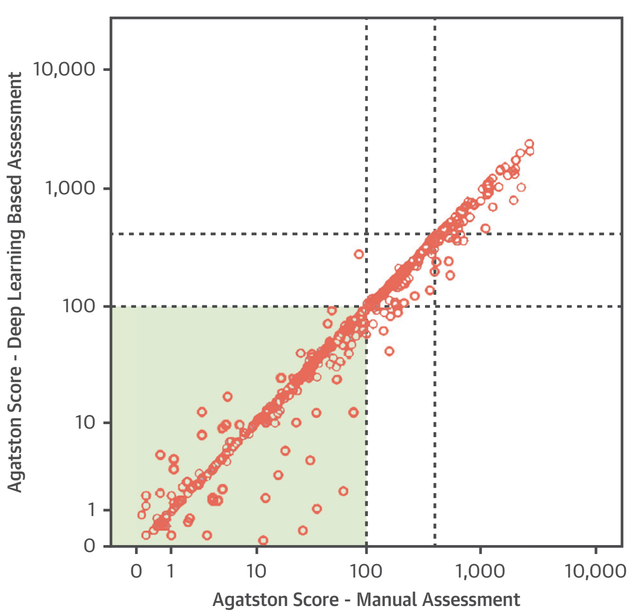
In total, 997 participants from the Robinsca trial who underwent low-dose electrocardiography-triggered cardiac CT (Somatom Flash, Siemens Healthineers) for calcium scoring were included between December 2015 and June 2016. The median age was 61.0 years (IQR: 11.0 years), and 54.4% were male. Detailed inclusion criteria have been published previously (5). All participants provided written informed consent, and the Robinsca study was approved by the Minister of Health after a positive recommendation by the Health Council, was supported by the European Research Council (294604), and is registered under NTR6471.

Standardized CT screening protocols with a tube voltage and current of 120 kVp and 80 reference mAs/rotation were applied in high-pitch spiral mode. Acquired images were reconstructed with filtered back projection, sharp reconstruction kernel (b35f) slice thickness, an increment of 3.0/1.5 mm, and a field of view of 250 mm.

CAC was identified, quantified using the Agatston score, and classified by a fully automated deep learning-based CAC scoring prototype (AVIEW CAC, Coreline Soft). The software was developed based on a 3-dimensional U-net architecture using non-enhanced cardiac CT scans acquired from multiple vendors and scanners. No training data were included in this current study. The original manual assessment of the Robinsca trial was used as the reference, and this was performed with dedicated software (Syngo.via CaSc, version VB10A, Siemens) by 2 readers with experience in more than 1,000 CAC CT scans.

CAC measurement (Agatston score) and risk categorization (0-99, 100-399, and ≥400) of the deep learning prototype were compared to the original manual assessment. A high agreement for detection was found between deep learning-based and manual scoring) (κ = 0.87; 95% CI: 0.85-0.89). Median Agatston score of all positive cases was 58.4 (IQR: 12.3-200.2) and 61.2 (IQR: 13.9-212.9) for deep learning-based and manual assessment respectively (intraclass correlation coefficient: 0.958; 95% CI: 0.951-0.964) (Figure 1). The reclassification rate was 2.0%, with a very high agreement (κ = 0.960;

FIGURE 1 Correlation Coronary Artery Calcification Scoring



Positive Agatston score for deep learning-based vs manual assessment. The dotted lines represent the risk category cutoff values 100 AU and 400 AU, with the majority of individuals below 100 **(shaded quadrant)**. Overall, a high correlation (intraclass correlation coefficient: 0.958) was found for positive Agatston scores.

95% CI: 0.943-0.997; $P < 0.001$); none of the cases shifted more than 1 category. The threshold for the initiation of preventive treatment in cardiovascular disease screening is an Agatston score ≥100 Agatston unit. This resulted in a false negative rate of 0.7%, a false positive rate of 0.1%, and a diagnostic accuracy of 99.2% for the initiation of preventive treatment.

The deep learning-based software for automatic CAC scoring performed excellently in a population-based screening setting to determine risk categorization in asymptomatic participants.

Future deep learning software that is able to assign a limited number of uncertain cases for manual human feedback could improve the calcium scoring process and outperform (a panel of) experienced readers that solely use manual scoring.

Limitations of the study were that the evaluation of the deep learning software was focused on a cardiac screening setting only, comprising the screening distribution of calcium and a highly standardized protocol to ensure high reproducibility and accurate risk stratification. Therefore, the generalizability of the current results to a clinical setting comprising a different population and variable scanner settings may be limited.

In conclusion, deep learning-based software for automatic CAC scoring can be used in a cardiovascular CT screening setting with high accuracy for risk categorization and the initiation of preventive treatment.

Marleen Vonder, PhD*
Sunyi Zheng, PhD
Monique D. Dorrius, MD, PhD
Carlijn M. van der Aalst, PhD
Harry J. de Koning, MD, PhD
Jaeyoun Yi, PhD
Donghoon Yu, MSc
Jan Willem C. Gratama, MD, PhD
Dirkjan Kuijpers, MD, PhD
Matthijs Oudkerk, MD, PhD

*University of Groningen
University Medical Center Groningen
Department of Epidemiology
PO Box 30.001
FA 40, 9700 RB
Groningen, the Netherlands
E-mail: m.vonder@umcg.nl

https://doi.org/10.1016/j.jcmg.2021.07.012

The Robinsca study was supported by an advanced grant of the European Research Council (294604). Drs Gratama, Kuijpers, and Oudkerk have received research support from Institute for Diagnostic Accuracy, Groningen, the Netherlands. Dr Yi and Yu are employees of Coreline Soft, Seoul, rep of Korea. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the Author Center.

REFERENCES

1. Silverman MG, Blaha MJ, Krumholz HM, et al. Impact of coronary artery calcium on coronary heart disease events in individuals at the extremes of traditional risk factor burden: the Multi-Ethnic Study of Atherosclerosis. *Eur Heart J*. 2014;35(33):2232–2241.

2. Grundy SM, Stone NJ, Bailey AL, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol. *J Am Coll Cardiol*. 2019;73:e285–e350.

3. Wang W, Wang H, Chen Q, et al. Coronary artery calcium score quantification using a deep-learning algorithm. *Clin Radiol*. 2020;75:237.e11–237.e16.

4. de Vos BD, Wolterink JM, Leiner T, de Jong PA, Lessmann N, Isgum I. Direct automatic coronary calcium scoring in cardiac and chest CT. *IEEE Trans Med Imaging*. 2019;38:2127–2138.

5. van der Aalst CM, Denissen SJAM, Vonder M, et al. Screening for cardiovascular disease risk using traditional risk factor assessment or coronary artery calcium scoring: the ROBINSCA trial. *Eur Heart J Cardiovasc Imaging*. 2020;21(11):1216–1224.

Predictors of Prosthetic Valve Regurgitation After Transcatheter Aortic Valve Implantation With ACURATE neo in the SCOPE I Trial

Adverse effect profiles of transcatheter heart valve (THV) systems for transcatheter aortic valve replacement (TAVR) differ. In the SCOPE I (Safety and Efficacy of the Symetis ACURATE Neo/TF Compared to the Edwards SAPIEN 3 Bioprosthesis; NCT03011346) trial, 739 patients were randomized to transfemoral TAVR with either the self-expanding ACURATE neo (NEO) or the balloon-expandable SAPIEN 3 (S3) THV (1). Details of the design have been published (2). Study approval was obtained from ethics committees at each site. NEO failed to meet noninferiority compared with S3 regarding a composite end point at 30 days driven by a higher rate of at least moderate prosthetic valve regurgitation (PVR) (2). Previous studies showed that moderate or more PVR is associated with an increased risk of adverse outcomes (3). The objective of the present analysis was to evaluate the impact of anatomical and procedural factors on the occurrence of relevant PVR after TAVR with the NEO device in the SCOPE I trial.

Baseline multidetector computed tomograms (MDCTs) were analyzed by an independent laboratory (Kerckhoff Heart Center, Bad Nauheim, Germany) using 3mensio software (Pie Medical Imaging). Cover index was defined as: 100 × ([THV diameter − perimeter-derived annulus diameter]/THV diameter). Calcium detection thresholds were based on the mean